

DroPNet: a web portal for integrated analysis of *Drosophila* protein–protein interaction networks

Yoan Renaud¹, Alexandre Baillif^{2,3}, Jean-Baptiste Perez^{2,3}, Marie Agier^{2,3},
Engelbert Mephu Nguifo^{2,3} and Vincent Mirouse^{1,*}

¹GReD Laboratory, Unité Mixte de Recherche Centre National de la Recherche Scientifique 6293, Clermont Université, Institut de la Santé Et de la Recherche Médicale U1103, Faculté de Médecine, Place Henri Dunant, 63000 Clermont-Ferrand, ²Clermont Université, Université Blaise Pascal, Laboratoire d'Informatique, de Modélisation et d'Optimisation des Systèmes, BP 10448, F-63000 Clermont-Ferrand and ³CNRS, UMR 6158, LIMOS, F-63173 Aubiere, France

Received January 31, 2012; Revised April 17, 2012; Accepted April 25, 2012

ABSTRACT

DroPNet (Drosophila Protein Network) is a *Drosophila*-dedicated web portal for generating and analyzing protein–protein interaction (PPI) networks. This platform integrates users' experimental data provided as one or two lists of genes with PPI data from *Drosophila* and other species. These experimental data can, for example, come from RNAi screens, for which this approach is known to be valuable. DroPNet, therefore, provides an essential basis for further biological analysis by linking functional and physical interactions and reinforcing the relevance of each. DroPNet focuses on the search of PPIs between genes of the entry list, and includes the possibility of searching for intermediate genes for which the corresponding protein indirectly links two entry data. It also offers multiple functions for editing the networks obtained, providing users with interactive possibilities to progressively improve and refine the results. This approach gives a global view of the studied process and makes it possible to highlight specific interactions that have so far been understudied. DroPNet is freely available at <http://dropnet.isima.fr>.

INTRODUCTION

Recent years have seen the development of large-scale functional approaches based on RNAi applications. These techniques are able to knock down the expression of a specific gene in cell culture or *in vivo*, to see whether it

has an effect on a specific biological event or function. This is especially true with the *Drosophila* model, where such screens are very easy to perform in cell culture as well as *in vivo* (1). In cell culture, RNAi can be very simply and efficiently induced without any need for transfection (2,3). *In vivo*, the RNAi, coupled to the ectopic expression system UAS/Gal4, makes it possible to knock down a selected gene in a specific population of cells during fly development or adult life (4). Moreover, this *in vivo* RNAi approach is now made possible at a whole-genome scale thanks to the development of *Drosophila* transgenic line collections that cover the entire genome (5,6). Thus, both in culture and *in vivo*, there is an exponentially growing number of reverse genetics screens being performed in *Drosophila* laboratories all over the world and covering all the domains of biology, from development to immunity or behavior.

The main advantage of the reverse genetics approach over classical genetic screening is that biologists immediately know exactly which genes are involved in the process they are studying. They, therefore, end up with a list of genes that may be short (less than 10) or to several hundred depending on the sensitivity and accuracy of their screen read-out. The problem at this point is how to most efficiently handle this list—an issue that raises several questions: is it possible to get a global view of the studied process? What is already known and what is really new in the data? What are the best candidate genes for in-depth research? What are the molecular links connecting all the members of this list?

One approach to addressing all these questions is to integrate the data from the functional screens with the data on protein–protein interactions (PPIs) referenced in available databases, which feature results from literature

*To whom correspondence should be addressed. Tel: +33 4 73 17 81 71; Fax: +33 4 73 27 61 32; Email: vincent.mirouse@u-clermont1.fr

and high-throughput proteomics screens. However, these approaches generate significant numbers of false positives, which ultimately means their results are underexploited. Based on the PPIs, it is possible to map protein interaction networks that can establish molecular links between proteins coded by genes that share functional links. Grouping proteins into a common network based on a shared combination of function and physical interaction should considerably increase relevance confidence of both. This is a concept that has been validated both statistically and experimentally by independent groups (7,8). However, although protein network analysis web servers exist, none of them has made the analysis task easy to perform. We, therefore, developed a web platform called DroPNet (*Drosophila Protein Network*) with two main goals: (i) to include options that improve the possibilities of the approach, and these options are unique to DropNet; (ii) to build an easy-to-use tool specifically dedicated to *Drosophila* (although integrating data from other species) to ensure that any biologist working on the *Drosophila* model can perform this integrative approach and gain positive results.

OVERVIEW

There are a number of applications available for drawing PPI networks, and each has its own aims and possibilities. Cytoscape is an application that offers the possibility of drawing networks from the user's own data (9). Many plug-ins are available for Cytoscape, and some of them, like APID, BioNetBuilder, MIMI or Droid plug-ins, allow users to search through public PPI data (including *Drosophila*'s data) using the user's list of genes and retrieve the interactions (10). However, Cytoscape is not a web server-based application. STRING is a database of known and predicted protein interactions. STRING also offers an online visualization tool that gives the possibility of viewing protein interactions referenced in the String database as a network (11). Similarly, a visualization tool called IM Browser (for Interaction Map Browser) developed by the Finley Lab is able to work with DroID data (12,13). PINA provides a non-redundant protein interaction dataset for six model organisms and includes a tool for construction, visualization and management of protein interaction networks (14,15). GeneMANIA is an other web-based PPI visualization tool that also integrates other kind of data and aim to predict gene function (16). We have a step further and developed functionalities that are specific to DroPNet. DroPNet focuses on the search for both direct and indirect PPI among genes given by the user. Moreover, this search can be done on a unique list of genes or between two lists. An average number of interactions obtained from 10 random genes lists of same size as user gene list is provided for each network generated, giving an estimation of the significance of the interaction number. Another important feature of DroPNet (called 'Upgrade') is a system of iterative searches based on manual validation or deletion of interactions and new exploration for intermediates genes.

DROPNET WEB PLATFORM

DroPNet is a web platform that allows users to visualize, manipulate and save PPI networks with novel features for analysis.

As DroPNet is geared specifically to *Drosophila* interactions, it can take data from the DroID database (13). Indeed, *Drosophila* PPI data can be found in many public databases, including Intact (17), Mint (18) and Biogrid (19), and they extend to cover many species based on the notion of ortholog genes. This is one of the factors that guided our decision to choose DroID, as it already recovers data from these different sources and species. DroID has also recently been updated with fresh datasets from large-scale co-affinity purification coupled to mass spectrometry analysis (20).

The DroPNet interface is relatively simple yet complete, and features two main pages. The first page comprises a form enabling users to fill out data on their genes of interest and certain user-selected parameters. The second page gives the search results output as a Java applet displaying the PPI network and containing additional features to improve the obtained graph.

Implementation

The DroPNet web platform runs on a Transtec 2300L Data Storage Server running Intel Xeon E5506 quad-core 2.13 GHz processors with 4 Mb of L2 cache and 8 Gb of RAM. The form parts are implementing using the framework Richfaces 3.3.2 from JBOSS. Richfaces is a framework that simplifies the using of AJAX with JSF 1.2 technology. The part of drawing the networks is a Java applet, built in with the JDK6 version. This applet uses an open-source and free library for drawing networks called Piccolo2D 1.2.1.

Data entries

The input screen of the DroPNet web platform is shown in Figure 1 (containing default parameter values).

To generate the PPI network, the user needs to list representative genes for each protein he/she wants to consider, and to check a number of parameter boxes.

Initially, the user has to provide one or two lists of genes for generating the network (Figure 1a). The major advantage of considering two lists is to allow the user to look for interactions that exclusively involve genes from one list and genes from the other list. This approach can be valuable if the user wants to analyze two lists of genes grouped by functionality to see whether they show interactions. For instance, two sets of genes sharing phenotype similarities could be analyzed using this option to look for PPIs creating physical links between the two sets. This choice option can be specified in a checklist on the input screen. List(s) of genes have to be entered using their CG number or FB number in a specific format (one number per line or inline list separated with comma or space) that is checked by the data entries web page. We chose not to use name as identifier as a gene can have many synonyms and people could misspell a name, making it time consuming to run input verifications. We also decided

Please enter your data

(d) Fill sample data: 1 2 3 4

Search interactions in:

☐ Only in A

☐ Between A and B (exclusive)

☐ Between A and B (inclusive)

Number of intermediate proteins: 2 (b)

Filter intermediate by number of interactions: 50

* Only interactions between list A and list B will be considered

** Interactions between list A and list B will be considered plus interactions between genes in A and between genes in B

In which sources:

Get a description of databases here.

Select All Drosophila Only

Unselect All Orthologs Only

☒ Curagen yeast two-hybrid

☐ Finlay Lab yeast two-hybrid

☐ HybriGenics yeast two-hybrid

☐ Other physical interactions

☒ DPIM co-AP/MS

☐ Human interologs

☐ Yeast interologs

☐ Worm interologs

Generate Network

Clear Data

Figure 1. Input screen of the DroPNet web server showing a predefined sample. (a) The two lists of genes to be filled in. (b) Drop-down lists to parameterize intermediate components. (c) List of the different available sources. (d) Sample Data.

against using protein IDs as there are often several isoforms for a given gene, which can produce misleading results.

Next, users have to specify a small set of parameters to define how they want to construct their network. In fact, the application looks for direct interactions between two of the given proteins, but sometimes two proteins can only be linked indirectly by one or more other proteins. These proteins are called intermediate components, and DroPNet offers the possibility of searching for them. It could be useful to highlight when two proteins interact indirectly via another protein that was not given as input data. This is a selectable option in the two drop-down lists on the input screen (Figure 1b).

'Number of intermediate proteins': users can choose to permit intermediate proteins (i.e. non-listed proteins that allow interactions by transitivity with the proteins entered). In the platform, to keep a reasonable calculation time and output easy-to-manipulate networks, number of intermediates is limited to 2. Moreover, tests have shown that with more than two intermediates, the biological meaning of these indirect interactions has little relevance and leads to inopportune extension of the networks.

'Filter intermediate by number of interactions': users can filter the intermediate proteins they want to consider in the network by using a maximum interaction criterion. For example, to filter out intermediate proteins that occur in more than 50 interactions, the user will choose 50 in the drop-down list. Several values between 25 and 500 can be chosen. The No Limit value means that the user does not want to apply any filter on intermediate proteins. This option is very useful for removing proteins with low binding specificity, such as chaperone proteins. The indirect links created by such intermediates is usually not relevant. We opted for this filter system rather than one based on a confidence score. First, not all the interactions referenced in databases come with a confidence score. Second, high scores are mainly due to the fact that an interaction has already been extensively studied.

Consequently, filtering on such scores would remove all the potentially interesting understudied interactions, which is exactly the reverse of one of the main goals of this application.

User can select different sources for interactions from different databases as classified in DroID (Figure 1c). Two methods are available for executing this task: the user can check each interaction database listed one by one, or else they can choose to select all databases or only both *Drosophila* and ortholog databases simply by clicking on the associated button(s). Hyperlinks are available to get a more detailed description of each database.

A detailed description of all the features of the web server is available in the Tutorial Page that is accessible in the banner of each page of the DroPNet platform. Note that the web platform can be tested with four predefined samples on the top left hand corner of the input page. If the user clicks one of the samples, it will automatically fill lists of genes and check a set of parameters. For instance, the first sample is real results from a whole genome RNAi screen on centriole duplication and maturation, a key organizing structure for microtubule cytoskeleton (21).

Results page

Once users have submitted their data and filled in the different parameters, they can generate the network. A Java applet will launch and the network will be drawn. Note that if the user wants to generate a network that turns out to be huge, he/she will be informed of its size (intermediate node number >70 and interaction number >500) before the network is generated. Moreover, the user can choose to continue generating it or to abort the process and return to the data entry page.

The results page is divided into three parts: a menu bar with a set of specific options (Figure 2d), the interactive network itself (Figure 2a) and an informative panel split into two boxes (Figure 2b and c). The main part of the page consists in the graph corresponding to the PPI network the user wants to obtain. Proteins are represented by colored round nodes, and interactions between two proteins are represented by colored dashed line edges. Each color of the different edges is associated with the database source corresponding to the interactions. If the user chooses to enter two distinct lists of genes in the input page, nodes will have different colors according to the list from which the gene comes. Moreover, a different color is used for intermediate nodes. Information on these different colors is available in the left part of the informative panel located at the bottom of the screen (Figure 2b). This information is followed by the number of interactions present in this network and the average number of interactions obtained from 10 randomly generated gene lists of the same size and based on the same parameters. These values give user an estimate of how their network is enriched in relevant interactions. It is worth noting that all the published RNAi screen results that we tested with DroPNet gave much bigger networks than random runs, whatever the parameters used, thus confirming the validity of the approach. For instance, we tested a series of parameters with sample 1 (Table 1). In all cases, the number

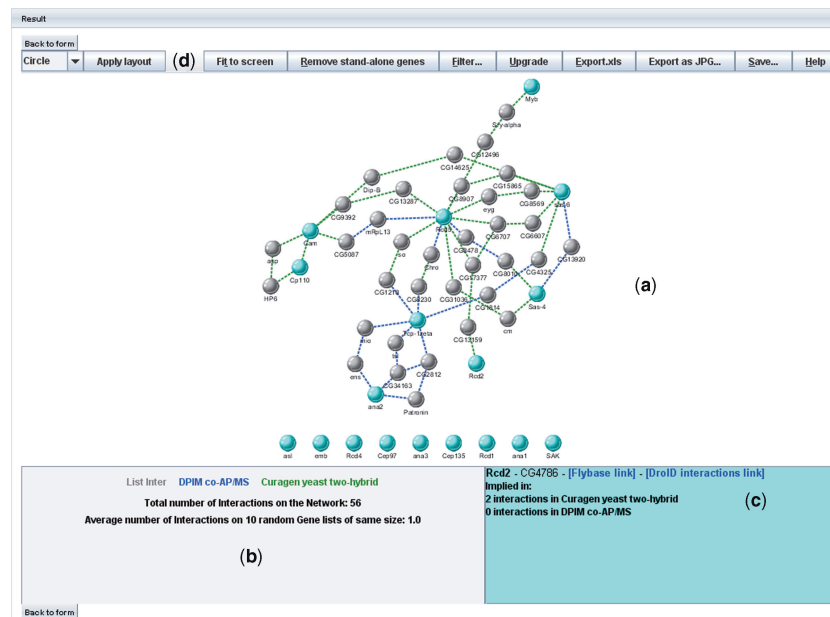


Figure 2. Results screen of the DroPNet web server. (a) Graph representing the interaction network targeted. (b) Legend panel containing graph-related information. (c) Detail panel on a considered edge or node. (d) Tools menu for editing or saving the network. The network shown here corresponds to Sample 1 and parameters indicated on Figure 1.

Table 1. Example of numbers of PPI found with different parameters applied on Sample 1

Selected databases	Zero intermediate	One intermediate (filter <100)	Two intermediates (filter <50)
Curagen and DPIM	1/0	8/0.4	56/2.4
Drosophila only	1/0	8/0.6	72/3.1
All	10/0.2	288/8.7	257/7.4

For each the first number corresponds to number of found interactions and the second one to the result average of 10 random sets of the same size and with the same parameters. The first selection (Curagen and DPIM) corresponds to two unbiased whole proteome screen.

of interactions found in the network is far bigger than the average obtained with random sets. It is worth noting that is also true when looking only to Curagen and DPIM data that have been obtained by whole genome proteomics approaches. Thus, integrating data from an unbiased whole genome RNAi screen with unbiased PPI data still lead to very significant results.

Users can easily interact with the network using mouse commands:

- A user can move a node by straightforward drag and drop. They can also move all the nodes by dragging the background.
- For a given node, a contextual menu (right-click) can be used to open a new table with a Flybase gene description page or to delete this node.
- By default, an interaction stands as ‘not validated’. By right-clicking an edge, the user can either validate it if he trusts this interaction or delete it if he thinks it is

not relevant. The validation of an interaction makes its edges becoming full-line. It is possible to undo this action by invalidating the interaction, which will return all the edges of the interaction to dashed line format. These options are important for improving network visualization, but they are also required for the ‘upgrade’ option described later.

- Left-clicking nodes or edges will bring up information on the item in the right part of the informative panel (Figure 2c). Clicking on a node brings up the gene name abbreviation with its associated CG Number and links to its corresponding web pages on both Flybase and DroID web sites. The panel also indicates the number of known interactions for each source involving the selected protein. Clicking on an edge will display the species where the interaction has been identified. The panel also gives a link to paper(s) where this interaction is mentioned and to original source information on the interaction.
- Scrolling the mouse wheel offers a zoom in/zoom out function.

The tool menu at the top of the results page offers more features for editing the network or saving the data and the resulting graph (Figure 2d). The first available option is the choice of layout that the user wants to apply for the generated interactions network. Instead of manipulating the graph, users can opt to apply predefined layout (available in a drop-down list) to their network designed to make analysis easier. However, our experience is that the different organic layouts give the best results. Three other functionalities are also designed to improve the network visualization display according to user choice. In practice, users can center and scale the network on-screen to make

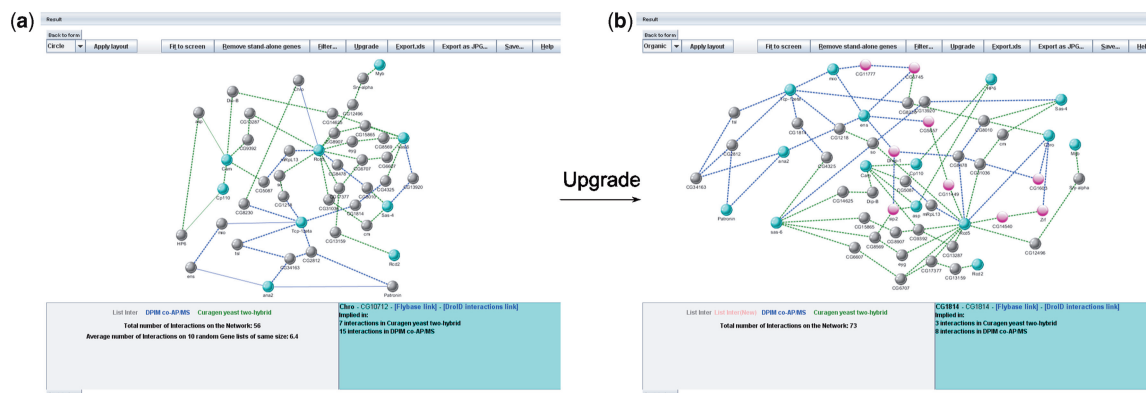


Figure 3. Results from an upgrade of the network shown in Figure 2. (a) Network where six interactions have been validated. (b) Resulting Network after upgrade. New intermediate nodes appear in pink.

it more readable, and also delete all stand-alone genes to get rid of all unnecessary information. The third visual display option consists in applying filters. Thus, users can control the visibility of both the interaction species they want to consider and the number of intermediate nodes in network, which should make specific analyses easier.

The tools menu also offers several saving features allowing the user to save their network in different ways. The user can save their network as a JPEG picture to keep a visual file, as well as export all data into an .xls file. The option enables the user to reuse the information obtained from the network, and table format data is often valuable for further analysis. Users are not required to log in to use DroPNet, but they can elect to create a log-in account to save their edited networks. Users will find it useful to be able to return later on to their network and find it at the point it was saved, and to be able to store different networks at different states of editing. It also guards against information being lost when the user exits the application.

Another important feature of DroPNet is that it proposes a system of iterative searches based on manual validation or deletion of interactions, called Upgrade. To use the Upgrade function, the user has to validate a set of network edges involving intermediate nodes. Note that for each time an internode is linked to a node of the initial list(s) by a validated edge, its associated gene will be added to the initial list. A new network is generated that integrates this new list of genes with the same parameters as the previous network. For instance, rapid overview on Flybase of the intermediate genes shown on Figure 2 allows to identify at least six microtubule binding proteins or proteins involved in meiosis and mitosis, which are processes highly dependent on microtubules. We, therefore, validated interactions with these proteins and upgrade the network and nine new interactors were identified (Figure 3). To our knowledge, this option aiming to progressively improve and extend networks based on user interaction validation is highly innovative within the PPI networks domain.

CONCLUSIONS

DroPNet is dedicated to the design of PPI networks based on a list of *Drosophila* genes. This approach has, for instance, been validated for interpretation of RNAi screens. The main outcomes that can be achieved using DroPNet use are (i) to get an overview of the molecular connectivity of user-entered genes in efforts to identify new functions for specific pathways or molecular complexes and (ii) to highlight understudied molecular interactions present in databases and emerge their potential functions. In practice, DroPNet was developed to manipulate lists of just a handful up to a couple of hundred genes, but it is clearly not a tool for the visualization of a complete interactome or of all the PPIs mobilizing a single protein. It has also been designed to allow major user editing interactivity so as to improve the obtained networks (searches for intermediates, filters, upgrade and saving functions, and more). Users may need to try several parameters and spend time optimize the results depending on the size of their entry list, their precise aims and the degree of connectivity between the genes.

Although initially designed for the analysis of RNAi screen results, DroPNet can be used for post-analysis in other systems biology approaches, such as co-expression data. Therefore, DroPNet will prove a valuable tool for the *Drosophila* community as a whole, and in the extending systems biology field.

ACKNOWLEDGEMENTS

We thank Jordan Raff for the permission of using data from their RNAi screen as a sample. We thank Finley lab for DroID database, which is really helpful for us. Finally, we are grateful to the fly labs who test DroPNet.

FUNDING

CNRS/INSERM ATIP/Avenir program and the Association pour la recherche sur le cancer (ARC) to Y.R. and V.M.; Lifegrid grant to A.B. and J.-B.P. Funding for open access charge: CNRS.

Conflict of interest statement. None declared.

REFERENCES

- Perrimon, N. and Mathey-Prevot, B. (2007) Applications of high-throughput RNA interference screens to problems in cell and developmental biology. *Genetics*, **175**, 7–16.
- Clemens, J.C., Worby, C.A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B.A. and Dixon, J.E. (2000) Use of double-stranded RNA interference in *Drosophila* cell lines to dissect signal transduction pathways. *Proc. Natl Acad. Sci. USA*, **97**, 6499–6503.
- Hammond, S.M., Bernstein, E., Beach, D. and Hannon, G.J. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature*, **404**, 293–296.
- Lam, G. and Thummel, C.S. (2000) Inducible expression of double-stranded RNA directs specific genetic interference in *Drosophila*. *Curr. Biol.*, **10**, 957–963.
- Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblaue, S. *et al.* (2007) A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*, **448**, 151–156.
- Ni, J.Q., Zhou, R., Czech, B., Liu, L.P., Holderbaum, L., Yang-Zhou, D., Shim, H.S., Tao, R., Handler, D., Karpowicz, P. *et al.* (2011) A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nat. Methods*, **8**, 405–407.
- Mummery-Widmer, J.L., Yamazaki, M., Stoeger, T., Novatchkova, M., Bhalerao, S., Chen, D., Dietzl, G., Dickson, B.J. and Knoblich, J.A. (2009) Genome-wide analysis of Notch signalling in *Drosophila* by transgenic RNAi. *Nature*, **458**, 987–992.
- Wang, L., Tu, Z. and Sun, F. (2009) A network-based integrative approach to prioritize reliable hits from multiple genome-wide RNAi screens in *Drosophila*. *BMC Genomics*, **10**, 220.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
- Prieto, C. and De Las Rivas, J. (2006) APID: Agile Protein Interaction DataAnalyzer. *Nucleic Acids Res.*, **34**, W298–W302.
- Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M. *et al.* (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.*, **37**, D412–D416.
- Pacifico, S., Liu, G., Guest, S., Parrish, J.R., Fotouhi, F. and Finley, R.L. (2006) A database and tool, IM Browser, for exploring and integrating emerging gene and protein interaction data for *Drosophila*. *BMC Bioinformatics*, **7**, 195.
- Yu, J., Pacifico, S., Liu, G. and Finley, R.L. (2008) DroID: the *Drosophila* Interactions Database, a comprehensive resource for annotated gene and protein interactions. *BMC Genomics*, **9**, 461.
- Wu, J., Vallenius, T., Ovaska, K., Westermarck, J., Mäkelä, T.P. and Hautaniemi, S. (2009) Integrated network analysis platform for protein-protein interactions. *Nat. Methods*, **6**, 75–77.
- Cowley, M.J., Pinese, M., Kassahn, K.S., Waddell, N., Pearson, J.V., Grimmond, S.M., Biankin, A.V., Hautaniemi, S. and Wu, J. (2012) PINA v2.0: mining interactome modules. *Nucleic Acids Res.*, **40**, D862–D865.
- Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., Franz, M., Grouios, C., Kazi, F., Lopes, C.T. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **38**, W214–W220.
- Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S., Vingron, M., Roechert, B., Roepstorff, P., Valencia, A. *et al.* (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res.*, **32**, D452–D455.
- Zanzoni, A., Montecchi-Palazzi, L., Quondam, M., Ausiello, G., Helmer-Citterich, M. and Cesareni, G. (2002) MINT: a Molecular INteraction database. *FEBS Lett.*, **513**, 135–140.
- Stark, C., Breitkreutz, B.J., Reguly, T., Boucher, L., Breitkreutz, A. and Tyers, M. (2006) BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.*, **34**, D535–D539.
- Guruharsha, K.G., Rual, J.F., Zhai, B., Mintseris, J., Vaidya, P., Vaidya, N., Beekman, C., Wong, C., Rhee, D.Y., Cenaj, O. *et al.* (2011) A protein complex network of *Drosophila melanogaster*. *Cell*, **147**, 690–703.
- Dobbelaere, J., Josué, F., Suijkerbuijk, S., Baum, B., Tapon, N. and Raff, J. (2008) A genome-wide RNAi screen to dissect centriole duplication and centrosome maturation in *Drosophila*. *PLoS Biol.*, **6**, e224.